Learning Mixtures of Gaussians in High Dimension

Mikhail Belkin

The Ohio State University, USA
mbelkin@cse.ohio-state.edu

The study of Gaussian mixture distributions goes back to the late 19th century, when they were introduced by Pearson. Gaussian Mixtures have since become one of the most popular tools for modeling and data analysis, extensively used in speech recognition and other fields, due, in part to their simple mathematical formulation. Yet their properties are still not well understood. Widely used algorithms, such as Expectation Maximization (EM) often fail even on simple generated data and their theoretical properties are often unclear.

In my talk I will discuss some theoretical aspects of the problem of learning Gaussian mixtures. In particular, I will discuss our recent result, which, in a certain sense, completes work on an active recent topic in theoretical computer science by establishing quite general conditions for polynomial learnability of mixtures in high dimension by using techniques from semi-algebraic geometry.

The talk is based on joint work with Kaushik Sinha.